Part-of-Speech-Tagging Project Discussion

Hongyu Zhang

Compare Results

Experiment 1 accuracy values: Bigram HMMs without smoothing.

{0.01: 0.7063655030800822, 0.05: 0.8275154004106776,

0.1: 0.9014373716632443, 0.25: 0.9373716632443532,

0.5: 0.9507186858316222, 0.75: 0.9640657084188912,

1.0: 0.9691991786447639}

Experiment 2 accuracy values: Trigram HMMs without smoothing.

{0.01: 0.35831622176591377, 0.05: 0.7002053388090349,

0.1: 0.7915811088295688, 0.25: 0.8726899383983573,

0.5: 0.9301848049281314, 0.75: 0.9363449691991786,

1.0: 0.9404517453798767}

Experiment 3 accuracy values: Bigram HMMs with smoothing.

{0.01: 0.7751540041067762, 0.05: 0.8809034907597536,

0.1: 0.9065708418891171, 0.25: 0.9332648870636551,

0.5: 0.9486652977412731, 0.75: 0.9640657084188912,

1.0: 0.9681724845995893}

Experiment 4 accuracy values: Trigram HMMs with smoothing.

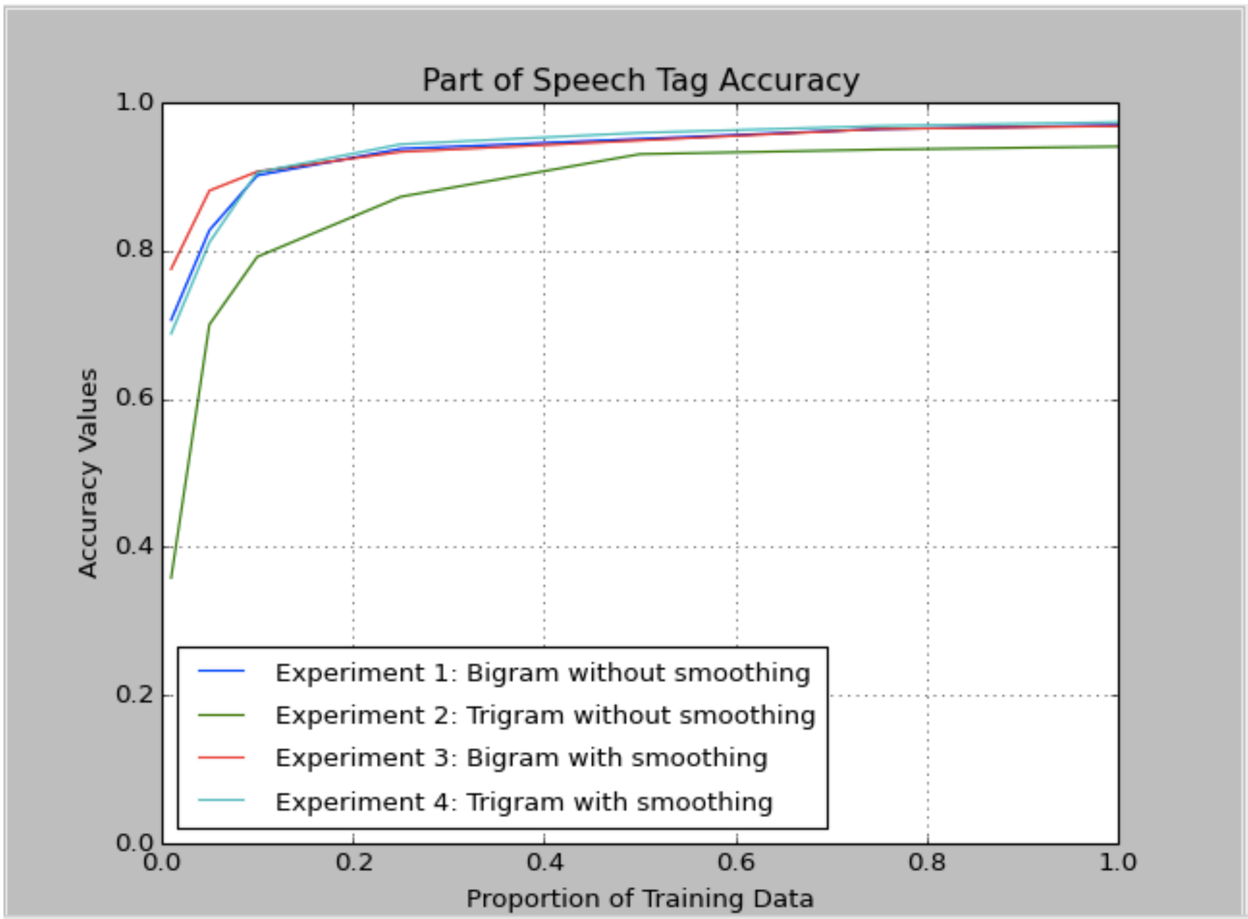
{0.01: 0.6878850102669405, 0.05: 0.811088295687885,

0.1: 0.9055441478439425, 0.25: 0.9435318275154004,

0.5: 0.9589322381930184, 0.75: 0.9681724845995893,

1.0: 0.973305954825462}

Plot those 4 results.



Discussion and Conclusion

Contrasting accuracies in a given experiment.

For each experiment, when the percentage of training data increases from 0.01 to 0.05 to 0.10 to 0.25, our corresponding accuracy values increases significantly. When the percentage of training data increases from 0.25 to 0.50 to 0.75 to 1.00, the increase in our corresponding accuracy values is not very significant.

For each of the experiment, we can see that as the percentage of training data increases, the accuracy values increase. For each experiment, the lowest percentage (1%) of training data used corresponds to the lowest accuracy values. And the highest percentage (100%) of training data used corresponds to the highest accuracy values.

When we have a small training data, there may be many new words that do not appear in the training data. Therefore, we update the emission matrix and give each new world a small possibility from each state. This will affect the emission matrix. Therefore, in Viterbi, when choosing tags based on probability, the accuracy may not be very high. However, when we increase the size of training data, our model will learn more words with their tags. And it will be less likely to have a lot of new words. Therefore, the accuracy may improve.

Contrasting accuracies across experiments.

When we have 1% or 5% training data, bigram with smoothing has the best accuracy, bigram without smoothing is the second, trigram with smoothing is the third, and trigram without smoothing has the lowest accuracy.

When we have 10%, 25%, 50%, 75% or 100% training data, trigram with smoothing has a little bit higher accuracy than others. Accuracy values for bigram with smoothing and bigram without smoothing are approximately the same. However, for trigram without smoothing, the accuracy values for each percentage is always the lowest one among all the experiments.

Conclusion

Based on our accuracy plot and our discussion, we can see that trigram models have a better training ability. When we increase the size of our training data, accuracies for trigram models improve a lot. This is because trigram models consider the preceding 2 words with tags instead of only 1 preceding word with tag (bigram). Therefore, trigram with smoothing has better accuracy than bigram with smoothing when we increase the size of training data.

When we calculate the transition probability, smoothing models considers $\lambda_1 * C(t_{i-1}, t_i) / C(t_{i-1}) + \lambda_0 * C(t_i) / NumTokens$ for trigram model and $\lambda_0 * C(t_i) / NumTokens$ for bigram models, while unsmoothing models do not consider these. Therefore, in general, smoothing models has better accuracy than unsmoothing models.

In accuracy graph, we see that for bigram, whether smoothing or not does not significantly affect the accuracy values when we have relatively large data (10%, 25%, 50%, 75%, 100%). However, for trigram, whether smoothing or not affects significantly about the accuracy values regardless of how large our data is. Trigram smoothing is always better than trigram unsmoothing. Trigram unsmoothing models do not have very accurate transition probabilities.